XIII. PROCESSING AND TRANSMISSION OF INFORMATION*

Prof. R. G. Gallager
Prof. F. C. Hennie III
Prof. I. M. Jacobs
Prof. R. S. Kennedy
Prof. A. M. Manders
Prof. B. Reiffen
Prof. C. E. Shannon
Prof. I. C. Stiglitz
Prof. J. M. Wozencraft
E. R. Berlekamp
D. G. Botha
H. Dym
P. M. Ebert
D. D. Falconer
E. Ferretti
G. D. Forney, Jr.
N. Gramenopoulos
J. L. Holsinger
K. Ikushima
Y. Iwadare
C. W. Niessen
R. Pilc
J. E. Savage
J. R. Sklar
K. D. Snow
R. N. Spann
W. R. Sutherland
M. G. Taylor
H. L. Yudkin

A. AN IMPROVED LOW-RATE ERROR BOUND FOR FIXED TIME-CONTINUOUS CHANNELS WITH MEMORY

1. Introduction

The author has presented[1] a random coding bound on probability of error for digital communication over the channel of Fig. XIII-1. As is usually the case with random coding bounds,[2,3] this bound can be shown to be quite poor under the conditions of low rate and high signal power. In fact, it is not difficult to show that the true exponent in the bound on probability of error differs from the random coding exponent by an arbitrarily large amount as the rate approaches zero and the signal power approaches infinity. This report presents an improved low-rate random coding bound based upon a slight generalization of recent work by Gallager[4] that overcomes this difficulty.

In this report all notation conforms with that used previously by the author.[1,5]

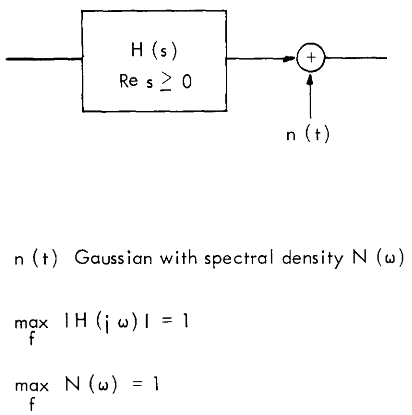

Fig. XIII-1. Gaussian channel with memory.

2. An Expurgated Random Coding Bound

Before proceeding with a derivation of the improved bound, it is important to consider briefly the reason for the inaccuracy of the random coding bound. In principle, the random coding bound is derived by considering a random ensemble of codes, each code containing M code words, such that the average of the energy of all of the code words in all of the codes satisfies a specified constraint. The probability of error, P_e, for each code

*This research was supported in part by the National Science Foundation (Grant G-16526), the National Institutes of Health (Grant MH-04737-03), the National Aeronautics and Space Administration (Grants NsG-334 and NsG-496); and in part by Purchase Order DDL BB-107 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the U. S. Air Force under Contract AF 19(628)-500.

is calculated, and the resulting values are averaged over the ensemble of codes to obtain $\overline{P_e}$. (In practice, it is the fact that the ensemble average may be performed first that makes this technique extremely useful.) Because of this averaging procedure, $\overline{P_e}$ is relatively large for an ensemble of codes containing a few codes with large P_e compared with the value that would be obtained if these codes were eliminated before averaging. (This is simply illustrated by considering the average of 10^{-1} and 10^{-10}.) An improved low-rate bound on P_e is derived here by expurgating the codes with high P_e from the ensemble used in the earlier derivation.

Consider the channel of Fig. XIII-1 and let $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_M$ be a set of code words for use with this channel. Then, given that $\underline{x}_j$ is transmitted, it has been shown[6] that

$$P_j(e) \leq \int_{\underline{Y}} P(\underline{y}|\underline{x}_j)^{1/2} \left[\sum_{\substack{k=1 \\ k\neq j}}^{M} P(\underline{y}|\underline{x}_k)^{1/2} \right] d\underline{y} \tag{1}$$

or, equivalently,

$$P_j(e) \leq \sum_{\substack{k=1 \\ k\neq j}}^{M} Q(\underline{x}_j, \underline{x}_k) \tag{2}$$

where

$$Q(\underline{x}_j, \underline{x}_k) = \int_{\underline{Y}} P(\underline{y}|\underline{x}_j)^{1/2} P(\underline{y}|\underline{x}_k)^{1/2} d\underline{y}.$$

From these equations it is clear that $P_j(e)$ is a function of each of the code words in the code. Thus, for a random ensemble of codes in which each code word is chosen with a probability measure $P(\underline{x})$, $P_j(e)$ will be a random variable, and it will be meaningful to discuss the probability that $P_j(e)$ is not less than some number A, that is, $\Pr\{P_j(e) \geq A\}$. To proceed further, a simple bound on this probability is required. Following Gallager,[4] let a function $\phi_j(\underline{x}_1, \ldots, \underline{x}_M)$ be defined as

$$\phi_j(\underline{x}_1, \ldots, \underline{x}_M) = \begin{cases} 1 & \text{if } P_j(e) \geq A \\ 0 & \text{if } P_j(e) < A \end{cases} \tag{3}$$

Then, with a bar used to indicate an average over the ensemble of codes, it follows directly that

$$\Pr\{P_j(e) \geq A\} = \overline{\phi_j(\underline{x}_1, \ldots, \underline{x}_M)}. \tag{4}$$

From Eq. 2 an obvious inequality is

$$\phi_j(\underline{x}_1, \ldots, \underline{x}_M) \leq A^{-s} \sum_{\substack{k=1 \\ k\neq j}}^{M} Q(\underline{x}_j, \underline{x}_k)^s \qquad 0 < s \leq 1, \tag{5}$$

since the right-hand side is always non-negative (for $A > 0$) and is not less than 1 when $P_j(e) \geq A$ and $0 < s \leq 1$. Thus

$$\Pr\left\{P_j(e) \geq A\right\} \leq A^{-s} \sum_{\substack{k=1 \\ k\neq j}}^{M} \overline{Q(\underline{x}_j, \underline{x}_k)^s} \qquad 0 < s \leq 1 \tag{6}$$

where

$$\overline{Q(\underline{x}_j, \underline{x}_k)^s} = \int_{\underline{X}} \int_{\underline{X}} \left[\int_{\underline{Y}} P(\underline{y}|\underline{x}_j)^{1/2}\, P(\underline{y}|\underline{x}_k)^{1/2}\, d\underline{y}\right]^s P(\underline{x}_j)\, P(\underline{x}_k)\, d\underline{x}_j d\underline{x}_k.$$

In this form it is clear that $\overline{Q(\underline{x}_j, \underline{x}_k)^s}$ is independent of j and k and therefore that Eq. 6 reduces to

$$\Pr\left\{P_j(e) \geq A\right\} \leq (M-1)\, A^{-s}\, \overline{Q(\underline{x}_j, \underline{x}_k)^s} \qquad 0 < s \leq 1. \tag{7}$$

At this point it is convenient to arbitrarily choose A to make the right-hand side of Eq. 7 equal to 1/2. Solving for the value of A gives

$$A = [2(M-1)]^\rho \left[\overline{Q(\underline{x}_j, \underline{x}_k)^{1/\rho}}\right]^\rho \qquad 0 < s \leq 1. \tag{8}$$

Now, let all code words in the ensemble for which $P_j(e) \geq A$ be expurgated. Then from Eq. 8 all remaining code words satisfy

$$P_j(e) < [2M]^\rho \left[\overline{Q(\underline{x}_j, \underline{x}_k)^{1/\rho}}\right]^\rho \qquad \rho \geq 1, \tag{9}$$

where $\rho = 1/s$. Furthermore, since $\Pr\left\{P_j(e) \geq A\right\} \leq 1/2$ it follows that the average number of code words, M', remaining in each code satisfies $M' \geq M/2$. Thus there exists at least one code in the expurgated set containing not less than M/2 code words and having a probability of error for each code word satisfying Eq. 9. By setting $e^{RT} = M/2$, it follows from this that there exists a code for which

$$P_e < 4^\rho\, e^{-TE^e(R,\rho)} \qquad \rho \geq 1, \tag{10}$$

where

$$E^e(R,\rho) = E_o^e(\rho) - \rho R$$

and

$$E_o^e(\rho) \triangleq \frac{-\rho}{T} \ln \overline{Q(\underline{x}_j, \underline{x}_k)^{1/\rho}}.$$

This bound will now be applied to the channel of Fig. XIII-1. As before,[1] let

$$P(\underline{y}|\underline{x}) = \prod_{i\in I} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y_i - \sqrt{\lambda_i}\ x_i)^2\right], \tag{11}$$

let $P(\underline{x})$ be chosen to be

$$P(\underline{x}) = \prod_{i\in I} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i}{\sigma_i}\right)^2\right], \tag{12}$$

and let the input power constraint be given by

$$\sum_{i\in I} \sigma_i^2 = ST. \tag{13}$$

(This form for $P(\underline{x})$ has been chosen primarily for mathematical expediency. Results obtained by Gallager,[4] however, indicate that it is indeed a meaningful choice from the standpoint of maximizing the resulting exponent.) Substituting Eqs. 12 and 13 in Eq. 10 yields

$$E_o^e(\rho) = \frac{-\rho}{T} \sum_{i\in I} \ln\left\{\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\frac{x_1^2 + x_2^2}{\sigma_i^2}\right\} \times\right.$$

$$\left.\left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{4}\left[(y-\sqrt{\lambda_i}\ x_1)^2 + (y-\sqrt{\lambda_i}\ x_2)^2\right]\right\}dy\right]^{1/\rho} dx_1 dx_2\right\}$$

which, after the integrals are evaluated, becomes

$$E_o^e(\rho,\underline{\sigma}) = \frac{\rho}{2T} \sum_{i\in I} \ln\left(1 + \frac{\lambda_i\sigma_i^2}{2\rho}\right) \qquad \rho \geq 1. \tag{14}$$

Here,

$$\underline{\sigma} = (\ldots, \sigma_j^2, \ldots, \sigma_\ell^2, \ldots).$$

For fixed R and T, maximization of $E_o^e(\rho) - \rho R$ over ρ, $\underline{\sigma}$, and the set I gives the desired bound. (In the maximization over ρ, it is assumed that T is large enough to make the exponential term in Eq. 10 the dominating factor.) Comparing Eqs. 10 and 14 with Eq. 10 given previously[7] shows a somewhat surprising similarity in the analytical expressions for the two bounds. Because of this, the maximization procedure used previously can be applied without change to this problem to yield the final result

$$P_e < 4^{\rho} e^{-TE_T(\rho)} \qquad \rho \geq 1. \tag{15}$$

Here,

$$E_T(\rho) = \frac{S}{4} B_T(2\rho-1)$$

$$R(\rho) = \frac{1}{2T} \sum_{i=0}^{N-1} \ln \frac{\lambda_i}{B_T(2\rho-1)} - \frac{S}{4} \frac{B_T(2\rho-1)}{\rho}$$

$$\frac{1}{B_T(\cdot)} = \frac{\dfrac{ST}{[1+(\cdot)]} + \sum\limits_{i=0}^{N-1} \lambda_i^{-1}}{N},$$

and N satisfies

$$\lambda_{N-1} > B_T(2\rho-1) \geq \lambda_N.$$

As before, a bound that is more readily evaluated can be derived by considering Eq. 15 for $T \to \infty$. The result, whose proof is too long to be presented here, is

$$P_e < 4^{\rho} e^{-TE(\rho)} \qquad \rho \geq 1, \tag{16}$$

where

$$E(\rho) = \frac{S}{4} B(2\rho-1)$$

$$R(\rho) = \int_W \ln \frac{|H(j\omega)|^2}{N(\omega)\, B(2\rho-1)}\, df - \frac{S}{4} \frac{B(2\rho-1)}{\rho}$$

$$\frac{1}{B(\cdot)} = \frac{\dfrac{S}{2[1+(\cdot)]} + \displaystyle\int_W \frac{N(\omega)}{|H(j\omega)|^2}\, df}{W}$$

$$W = \left\{ +f: \frac{|H(j\omega)|^2}{N(\omega)} > B(2\rho-1) \right\}.$$

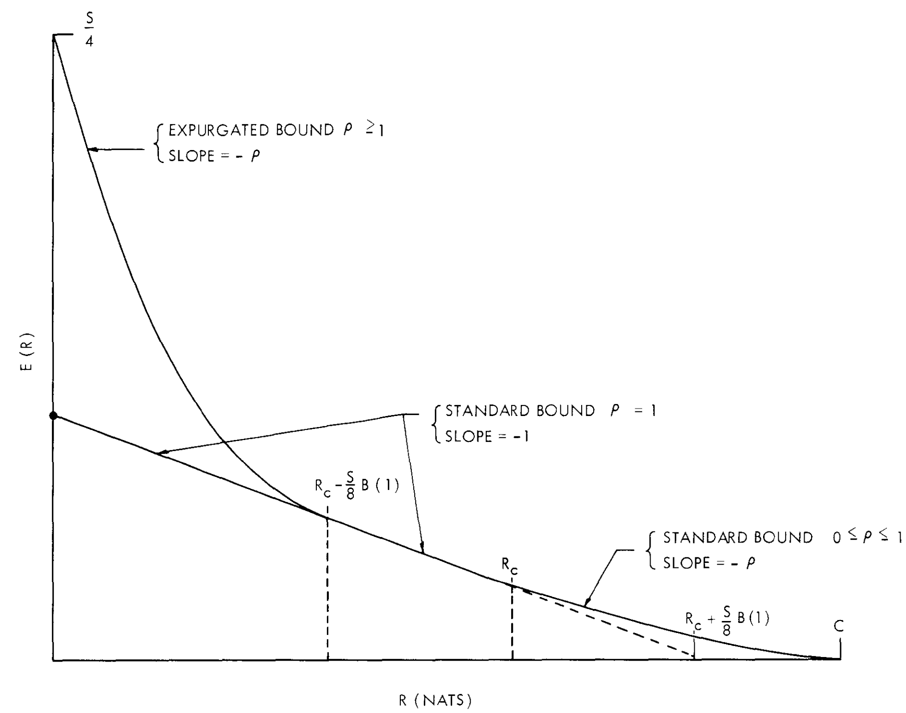

Fig. XIII-2. Error exponents for channel of Fig. XIII-1.

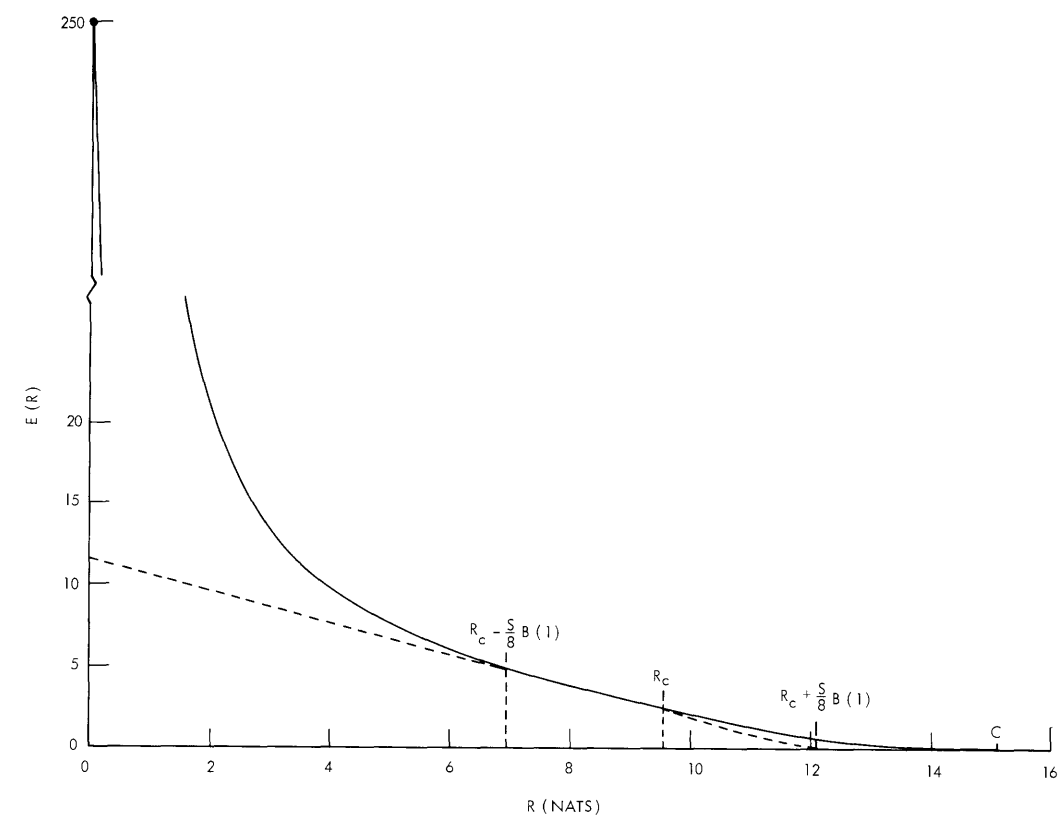

Fig. XIII-3. Error exponents for the channel of Fig. XIII-1 with $H(j\omega) = \left[1+j\,\frac{\omega}{2\pi}\right]^{-1}$, $N(\omega) = \frac{1}{2\pi}$, and $S = 10^3$.

Figure XIII-2 presents the pertinent characteristics of this bound and relates it to the previous random coding bound. Figure XIII-3 presents these bounds for a channel with $H(j\omega) = \left[1+j\frac{\omega}{2\pi}\right]^{-1}$, $N(\omega) = (2\pi)^{-1}$, and $S = 10^3$ and demonstrates the significant improvement of the expurgated bound.

J. L. Holsinger

References

1. J. L. Holsinger, An error bound for fixed time-continuous channels with memory, Quarterly Progress Report No. 72, Research Laboratory of Electronics, M.I.T., January 15, 1964, pp. 198-207.

2. C. E. Shannon, Probability of error for optimal codes in a Gaussian channel, Bell System Tech. J. 38, 611-656 (1959).

3. R. M. Fano, Transmission of Information (The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961), Chapter 9.

4. R. G. Gallager, A Simple Derivation of the Coding Theorem and Some Applications, unpublished manuscript.

5. J. L. Holsinger, Vector representation of time-continuous channels with memory, Quarterly Progress Report No. 71, Research Laboratory of Electronics, M.I.T., October 15, 1963, pp. 193-202.

6. J. L. Holsinger, An error bound for fixed time-continuous channels with memory, op. cit., page 200, Eq. 3.

7. Ibid., page 202.

B. LABORATORY SIMULATION OF A TIME-VARIANT MULTIPATH CHANNEL

A laboratory model of a time-variant multipath channel has been designed and constructed which provides delay-frequency spread products up to 1.2.

1. Motivation

In view of the mathematical complexity of analyzing the performance of communication systems operating over time-variant multipath channels, a laboratory model of such a channel has been built to aid in experimental investigations of interesting communication systems. The advantages of using a laboratory model rather than the real life channel are accessibility, convenience, lower cost, and less equipment complexity. A primary goal was to provide a channel that would exhibit as many real life channel problems as possible and also have a reasonable physical size. The model is not a simulation of any particular real channel; however, it does possess properties that are grossly similar to such channels as ionospheric scattering or the orbiting dipole channel.

2. Channel Parameters

In order for the model to have a long multipath delay, the wave propagation velocity in the channel must be relatively slow. Other required channel characteristics, such as adjustable multipath structure, low insertion loss, bandwidth, appreciable frequency spread, and a large ratio of bandwidth to frequency spread dictate the use of acoustic waves in water. Since the velocity of sound in water is approximately 1500 meters per second, a transmission path of 6 meters implies a time delay of approximately 4 msec. A simple way to provide both an adjustable multipath structure and time variations is to scatter and reflect sound waves from air bubbles in the water. This method as used in the channel model is shown in Fig. XIII-4.

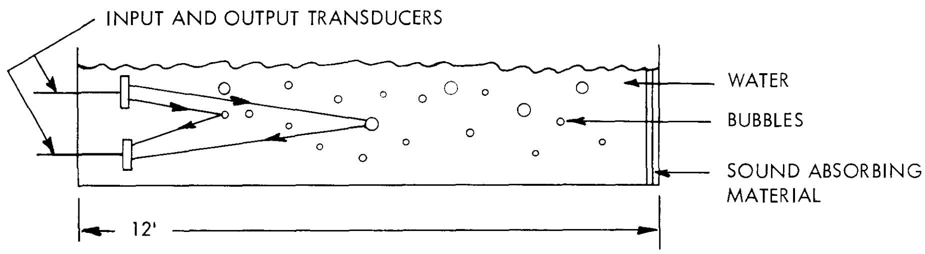

Fig. XIII-4. Channel model.

It is clear that if a short pulse is transmitted from the input transducer, similar pulses will be received from the output transducer at time delays corresponding to the bubble locations in the channel. For the 12 foot channel, the time delay between the first received pulse and the last received pulse is approximately 4 msec.

The input and output from the channel is through two barium titanate piezoelectric crystals cut to have a resonant frequency of 1 mc. May[1] has shown that a piezoelectric crystal operating near resonance has an equivalent circuit of a series RLC circuit in which R, and hence $\frac{1}{Q}$, increases when the mechanical loading on the crystal is increased. The desire for a wideband channel implies the need for low Q transducers operating at a high resonant frequency. With a conflicting factor of increasing attenuation of sound waves in water with increasing frequency, a reasonable compromise was a center frequency of 1 mc which yields an attenuation of only 0.1 db per yard and a channel bandwidth of 80 kc. In contrast, at 10 mc the attenuation is 10 db per yard.[2] The total insertion loss of the channel, measured as the ratio of average output voltage to sine-wave input voltage, and which includes attenuation, scattering from the air bubbles, and mismatches in coupling between the crystals and water, is −35 db.

3. Time Variations

The bubbles from which the sound waves in the channel are reflected are moving with some distribution of velocities and are within the channel for a length of time that also varies. At any one time, a large number of bubbles exists in the channel so that the channel output r(t) to any input x(t) will be a superposition of many waveforms similar to x(t) but differing in amplitude, phase, frequency, and duration. Specifically, if $x(t) = \cos \omega_o t$, we may write $r(t) = V(t) \cos (\omega_o t + \phi(t))$ or $r(t) = A(t) \cos \omega_o t + B(t) \sin \omega_o t$. At some time t_o the amplitude of $A(t_o)$ and $B(t_o)$ is due to scattering from a large number of bubbles, and by applying the Central Limit Theorem, the probability density of the random variables $A(t_o)$ and $B(t_o)$ tends to be Gaussian. Since $A(t_o)$ and $B(t_o)$ are also uncorrelated, it can be shown that the joint probability density of V(t) and $\phi(t)$ is

$$p_{V,\phi}(a,\theta) = \frac{a\, e^{-a^2/2a_o^2}}{2\pi a_o^2} \qquad \begin{array}{l} 0 \leq \theta \leq 2\pi \\ 0 \leq a < \infty. \end{array}$$

By integrating out θ,

$$\text{Prob}\,(V(t) > a) = e^{-a/a_o}.$$

Thus the probability that the received envelope V(t) exceeds some threshold is an exponential. This predicted exponential behavior of the envelope distribution function has been experimentally verified.

Finding an analytical expression for the power density function of r(t) is quite difficult, but it has been measured with a spectrum analyzer and a true rms meter. It is approximately Gaussian shaped, centered on 1 mc, with a spectral width of approximately 300 cps. Thus the maximum time delay-frequency spread product for the channel is approximately 1.2.

Present investigation of the channel concerns the statistical stationariness of the bandwidth, insertion loss, multipath structure, power density spectrum, and amplitude distribution.

K. D. Snow

References

1. J. E. May, Delay line equivalent circuits, J. Acoust. Soc. Am. 26, 347 (1954).
2. L. Beranek, Acoustical Measurement (John Wiley and Sons, Inc., New York, 1949).

C. AN ERROR BOUND FOR GAUSSIAN SIGNALS IN GAUSSIAN NOISE

In this report we use Gallager's bounding technique[1] to derive an upper bound on the probability of error for independent Gaussian signals in additive white Gaussian noise.

The results may be applied to the study of transmission of orthogonal signals through media in which the Gauss-Gauss model is reasonable (for example, fast Rayleigh fading or orbital dipole scattering).

Special cases of the result agree with previously known bounds for error probability of orthogonal signal transmission in additive white noise,[2] and of binary signalling with post detection diversity combining in the presence of Rayleigh fading,[3] and for the capacity of the infinite bandwidth Gaussian signals in a Gaussian noise channel.[4]

Suppose that we have a channel and a set of M messages $\{m_i\}$, $i = 1, 2, \dots, M$, with a corresponding set of signals such that, when the signal corresponding to message m_j is transmitted, the channel output is a set of M independent, n-dimensional vectors $\{\underline{y}_k\}$, $k = 1, 2, \dots, M$, whose components are statistically independent Gaussian random variables. Let y_{ki} be the i^{th} component of the k^{th} vector. We assume that

$$E[y_{ki}] = \begin{cases} 0; & k \neq j \\ a_i\sqrt{E_i}; & k = j \end{cases} \tag{1}$$

and

$$\sigma^2[y_{ki}] = \begin{cases} N_o/2; & k \neq j \\ \gamma_i E_i + N_o/2; & k = j, \end{cases} \tag{2}$$

where $\gamma_i \geq 0$ and $a_i^2 + \gamma_i \leq 1$. Let $\underline{y} = (\underline{y}_1, \underline{y}_2, \dots, \underline{y}_M) = (y_{11}, y_{12}, \dots y_{1n}, y_{21}, \dots y_{2n}, \dots y_{Mn})$. Then the conditional density function on $\underline{y}$ is given by

$$p(\underline{y}|m_j) = \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(\gamma_i E_i + N_o/2)}} \exp\left[-\frac{(y_{ji} - a_i\sqrt{E_i})^2}{2(\gamma_i E_i + N_o/2)}\right]\right] \prod_{k\neq j} \prod_{\ell=1}^{n} \frac{1}{\sqrt{\pi N_o}} \exp\left[-\frac{y_{k\ell}^2}{N_o}\right]. \tag{3}$$

For convenience we shall write Eq. 3 as

$$p(\underline{y}|m_j) = p_o(\underline{y}) \prod_{i=1}^{n} \frac{N_o/2}{E_i + N_o/2} f(y_{ji}) \exp\left[-\frac{a_i^2 E_i}{2(\gamma_i E_i + N_o/2)}\right], \tag{4}$$

where

$$p_o(\underline{y}) = \prod_{k=1}^{M} \prod_{\ell=1}^{n} \frac{1}{\sqrt{\pi N_o}} \exp\left[-\frac{y_{k\ell}^2}{N_o}\right] \tag{5}$$

is the probability density on the output arising from additive noise alone, and

$$f(y_{ji}) = \exp\left[\frac{y_{ji}^2}{2} \frac{\gamma_i E_i}{\frac{N_o}{2}(\gamma_i E_i + N_o/2)}\right] \exp\left[\frac{a_i\sqrt{E_i}\; y_{ji}}{\gamma_i E_i + N_o/2}\right]. \tag{6}$$

If we take n to be even and

$$\gamma_{2\ell-1} = \gamma_{2\ell}; \qquad \ell = 1, 2, \ldots, n/2,$$

we may view the channel as one in which $n/2$ diversity paths are available and for each message one of M orthogonal signals is transmitted over each diversity path. Each path is affected by independent Rayleigh fading. In this case E_i may be taken to be the energy transmitted on the i^{th} path, $a_i\sqrt{E_i}$ to be the amplitude of the "specular" component on the i^{th} path, and $\gamma_i E_i$ to be the variance of the random component on the i^{th} path.

Now, by symmetry, we have the fact that the average probability of error, $P(\epsilon)$, for the set of signals with a maximum-likelihood decision rule will be equal to the probability of error that is conditional on message 1 having been transmitted, $P(\epsilon|m_1)$. We compute an upper bound on this quantity, using an adaptation of a technique used by Gallager[1] to derive a coding theorem for discrete memoryless channels. We have

$$P(\epsilon) = P(\epsilon|m_1) = \int_{\underline{y}} p(\underline{y}|m_1)\, C_{m_1}(\underline{y})\, d\underline{y}, \tag{7}$$

where

$$C_{m_1}(\underline{y}) = \begin{cases} 1; & \text{if } p(\underline{y}|m_j) \geq p(\underline{y}|m_1) \text{ for any } j \neq 1 \\ 0; & \text{otherwise.} \end{cases} \tag{8}$$

Then for $\rho \geq 0$ we have

$$C_{m_1}(\underline{y}) \leq \frac{\left(\sum\limits_{j=2}^{M} p(\underline{y}|m_j)^{1/(1+\rho)}\right)^{\rho}}{p(\underline{y}|m_1)^{\rho/(1+\rho)}}. \tag{9}$$

This inequality follows from the fact that the right-hand side is always positive and at least one term will exceed 1 when $C_{m_1}(\underline{y}) = 1$. From Eqs. 4-6 we have

$$C_{m_1}(\underline{y}) \leq \left[\prod_{\ell=1}^{n} f(y_{1\ell})^{-\rho/(1+\rho)}\right]\left[\sum_{j=2}^{M} \prod_{i=1}^{n} f(y_{ji})^{1/(1+\rho)}\right]^{\rho}. \tag{10}$$

Substituting Eq. 10 in Eq. 7 and using the independence of the components of $\underline{y}$, we have

$$P(\epsilon) \leqslant \left\{\prod_{k=1}^{n} \overline{f(y_{1k})^{-\rho/(1+\rho)}}\right\} \left[\overline{\sum_{j=2}^{M} \prod_{i=1}^{n} f(y_{ji})^{1/(1+\rho)}}\right]^{\rho}, \tag{11}$$

where the bars indicate averaging with respect to $p(\underline{y}|m_1)$.

For $\rho \leqslant 1$, $X \geqslant 0$, X^{ρ} is a convex function with

$$\overline{X^{\rho}} \leqslant \overline{X}^{\rho}. \tag{12}$$

Thus

$$P(\epsilon) \leqslant \left[\prod_{k=1}^{n} \overline{f(y_{1k})^{-\rho/(1+\rho)}}\right] \left[\sum_{j=2}^{M} \prod_{i=1}^{n} \overline{f(y_{ji})^{1/(1+\rho)}}\right]^{\rho}$$

$$= (M-1)^{\rho} \left[\prod_{k=1}^{n} \overline{f(y_{1k})^{-\rho/(1+\rho)}}\right] \left[\prod_{i=1}^{n} \overline{f(y_{2i})^{1/(1+\rho)}}\right]^{\rho}, \tag{13}$$

where we have again used the independence of the components of $\underline{y}$. The computation of the averages in Eq. 13 may be performed by completing squares in the exponents of the integrands. Performing the required integration, we obtain

$$P(\epsilon) \leqslant \exp[-(E_1(\rho)+E_2(\rho)-\rho R^{*})], \tag{14}$$

where

$$E_1(\rho) = -\frac{1}{2} \sum_{i=1}^{n} \left(\ln \frac{1+\rho}{(1+\gamma_i\beta_i)\rho+1} + \rho \ln \frac{(1+\rho)(1+\gamma_i\beta_i)}{(1+\gamma_i\beta_i)\rho+1}\right) \tag{15}$$

$$E_2(\rho) = \frac{a_i^2\beta_i\rho}{2[(1+\gamma_i\beta_i)\rho+1]} \tag{16}$$

and we have defined $R^{*} = \ln(M-1)$ and $\beta_i = 2E_i/N_o$. Clearly,

$$E_2(\rho) \geqslant 0; \qquad 0 < \rho \leqslant 1, \ E_i \geqslant 0. \tag{17}$$

Using the inequality

$$-\ln x \geqslant 1 - x,$$

we find

$$E_1(\rho) \geqslant 0; \qquad 0 < \rho \leqslant 1, \ E_i \geqslant 0. \tag{18}$$

Differentiating, we also find

$$\frac{\partial^2 E_1(\rho)}{\partial \rho^2} \leq 0 \quad \text{and} \quad \frac{\partial^2 E_2(\rho)}{\partial \rho^2} \leq 0. \tag{19}$$

Thus the bound of Eq. 14 is minimized for fixed R^* and (β_i) by choosing ρ such that

$$R^* = \frac{\partial E_1(\rho)}{\partial \rho} + \frac{\partial E_2(\rho)}{\partial \rho}. \tag{20}$$

We then obtain the parametric bound

$$P(\epsilon) \leq \exp\left[\frac{1}{2} \sum_{i=1}^{n} \left(\frac{\gamma_i \beta_i \rho}{(1+\gamma_i\beta_i)\rho + 1} + \ln \frac{1+\rho}{(1+\gamma_i\beta_i)\rho + 1} - \frac{\alpha_i\beta_i(1+\gamma_i\beta_i)\,\rho^2}{(1+\gamma_i\beta_i)\rho + 1}\right)\right]$$

$$R^* = \frac{1}{2} \sum_{i=1}^{n} \left[\frac{\gamma_i\beta_i}{(1+\gamma_i\beta_i)\rho + 1} - \ln \frac{(1+\rho)(1+\gamma_i\beta_i)}{(1+\gamma_i\beta_i)\rho + 1} + \frac{\alpha_i\beta_i}{[(1+\gamma_i\beta_i)\rho + 1]^2}\right] \tag{21}$$

for $0 < \rho \leq 1$ and $R^* \geq \left.\frac{\partial E_1(\rho)}{\partial \rho} + \frac{\partial E_2(\rho)}{\partial \rho}\right|_{\rho=1}$. For $R^* < \left.\frac{\partial E_1(\rho)}{\partial \rho} + \frac{\partial E_2(\rho)}{\partial \rho}\right|_{\rho=1}$, we use the bound of Eq. 14 with $\rho = 1$. Calling the right-hand side of Eq. 21 $\exp[-E(R^*)]$, we have the result that

$$\frac{dE}{dR^*} = -\rho. \tag{22}$$

The optimization of the general bound with respect to the β_i is not readily accomplished analytically.

We now proceed to the investigation of two simple special cases.

In the case of orthogonal signals in white noise, we set

$$n = 1, \quad \alpha_1 = 1, \quad \gamma_1 = 0$$

and obtain

$$E_1(\rho) = 0$$

$$E_2(\rho) = \frac{ST}{N_o} \frac{\rho}{1+\rho}, \tag{23}$$

where we have defined $S = \frac{E}{T}$. Letting $R = \frac{R^*}{T}$ and $C = \frac{S}{N_o}$ and solving Eq. 21, we obtain the bound

$$P(\epsilon) \leq \exp[-CTE(R)], \tag{24}$$

where

$$E(R) = \begin{cases} \left(1 - \sqrt{\frac{R}{C}}\right)^2; & \frac{1}{4} \leq \frac{R}{C} \leq 1 \\ \frac{1}{2} - \frac{R}{C}; & 0 \leq \frac{R}{C} \leq \frac{1}{4}. \end{cases} \tag{25}$$

This bound has the same exponent as that given by Fano.[2]

In a special case of scatter channels with no specular component, we set

$$\alpha_i = 0; \quad \gamma_i = 1; \quad \beta_i = \beta = \frac{2E}{nN_o}; \qquad i = 1, 2, \dots n,$$

where E is the total transmitted energy. Equations 14 and 21 then give the bound

$$\begin{aligned} P(\epsilon) &\leq \exp\left[\frac{n}{2}\left(\frac{\beta\rho}{(1+\beta)\rho + 1} + \ln \frac{1+\rho}{(1+\beta)\rho + 1}\right)\right] \\ R^* &= \frac{n}{2}\left[\frac{\beta}{(1+\beta)\rho + 1} - \ln \frac{(1+\rho)(1+\beta)}{(1+\beta)\rho + 1}\right] \end{aligned} \tag{26}$$

for

$$0 < \rho \leq 1 \quad \text{and} \quad R^* \geq \frac{n}{2}\left[\frac{\beta}{2+\beta} - \ln \frac{2(1+\beta)}{2+\beta}\right],$$

and

$$P(\epsilon) \leq \exp\left[-\frac{n}{2} \ln \frac{(2+\beta)^2}{4(1+\beta)} + R^*\right] \tag{27}$$

for

$$R^* < \frac{n}{2}\left[\frac{\beta}{2+\beta} - \ln \frac{2(1+\beta)}{2+\beta}\right].$$

For fixed E/N_o the bound of Eq. 27 may be optimized with respect to n. Pierce[3] has optimized an exact expression for $P(\epsilon)$ in this case and found that, for large E/N_o, n should be picked such that

$$\beta \approx 3.$$

Substituting this value in Eq. 27 gives

$$P(\epsilon) \leq \exp[-0.149E/N_o + R^*], \tag{28}$$

which agrees with the exponent found by Pierce.

Taking lim $\rho \to 0$ in Eq. 26 gives the largest rate R^* for which the exponent in the bound on $P(\epsilon)$ is negative. This value of R^*/T may be taken to be a lower bound on the capacity of the channel under consideration. An evaluation of the limit gives

$$C = \frac{E}{TN_o}\left[1 - \frac{\ln(1+\beta)}{\beta}\right]. \tag{29}$$

Thus

$$\lim_{\beta\to\infty} C = S/N_o, \tag{30}$$

a result previously found by Jacobs.[4] Observe that for vanishing error probability for rates close to C Eq. 26 requires that $n \to \infty$.

The author wishes to acknowledge discussions with Professor Barney Reiffen.

H. L. Yudkin

References

1. R. G. Gallager, A simple derivation of the coding theorem, Quarterly Progress Report No. 69, Research Laboratory of Electronics, M.I.T., April 15, 1963, pp. 154-157.

2. R. M. Fano, Transmission of Information (The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961).

3. J. N. Pierce, Theoretical limitations on frequency and time diversity for fading binary transmissions, IRE Trans., Vol. CS-9, No. 2, pp. 186-187, June 1961.

4. Ira Jacobs, The asymptotic behavior of incoherent M-ary communications systems, Proc. IRE 51, 251-252 (January 1963).

D. A BOUND ON THE DISTRIBUTION OF COMPUTATION FOR SEQUENTIAL DECODING – AN EXAMPLE

1. Introduction

Recently, bounds have been obtained on the distribution of computation for the Wozencraft Sequential Decoding Algorithm, modified according to Gallager.[1] The bounds

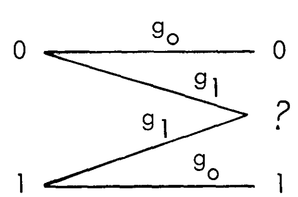

Fig. XIII-5. Binary Erasure Channel.

were derived for discrete, memoryless channels, uniform at the output. To illustrate the approach to the bounding, the same techniques are applied here to a particular algorithm for the Binary Erasure Channel (see Fig. XIII-5).

The results obtained for the Binary Erasure Channel as stated in Theorem 2, Eq. 16, and Fig. XIII-8 are analogous to the results for the more general channels.

2. The Algorithm

Before we define the algorithm, let us discuss the encoding process. We assume that a source supplies the encoder with a stream of equiprobable binary digits. This stream is encoded by using what is called a tree code (see Fig. XIII-6). If the first digit of the source stream is 0 (1) the upper (lower) branch is transmitted. The second and later source digits select branches according to the same rule. Since ℓ channel digits are transmitted for every source digit, the channel rate is $R = \frac{1}{\ell}$.

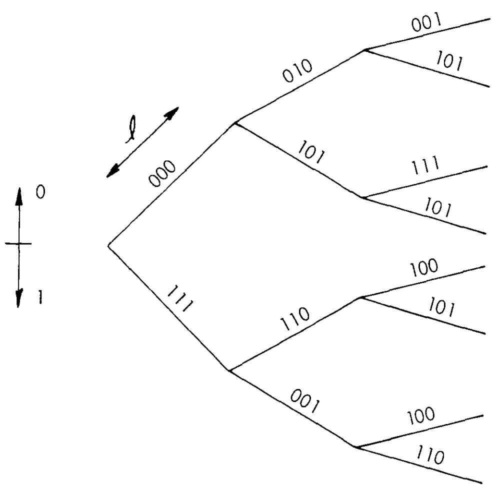

Fig. XIII-6. Tree code.

We note that the first source digit partitions the tree into two subsets and we call the subset containing the transmitted path the correct subset; the other is called the incorrect subset. The object of our decoding algorithm is to estimate each of the source digits by using the received data. The first source digit is decoded by determining which subset is the correct subset. Similarly, the second digit is decoded by determining the correct subset, given the first branch of that subset. Succeeding digits are decoded in the same manner. We assume that for each digit decoded, no more than S successive stages of the tree code are used. For this reason, S is called the decoding constraint length.

We now define a procedure for decoding the first source digit. It is related to the algorithm that is eventually used and it serves to introduce this second algorithm.

Assume that the decoder consists of two identical machines, each working on a subset. Each machine compares the first branch of its subset with the first ℓ digits of the received sequence. If any two unerased digits disagree, the tested branch is incorrect. The corresponding machine stops. The other machine is working on the correct subset and it must find its hypothesis branch acceptable. If both branches at the first stage are acceptable, the machines test the branches at the second stage. Paths at the second stage which disagree with the received sequence in unerased positions are also discarded. Each machine continues to operate until its subset is completely discarded or until it has searched all branches up to and including the S^{th} or last stage.

We assume that the machines operate independently and that they perform the same number of computations per unit time. Thus, machine no. 1 may be searching branches at the tenth stage while machine no. 2 is searching at the sixth stage. Let the machines do C_1 and C_2 computations, respectively, in searching through their subsets.

It should be clear that the procedure as defined can be improved. For example, if machine no. 1 eliminates its subset, then it is immediately obvious that machine no. 2 is working on the correct subset. Hence, the process can be terminated whenever a subset is eliminated. We now accept this modified algorithm as our decoding algorithm for the first digit. Succeeding digits are decoded in the same fashion by using that portion of the tree stemming from accepted branches.

The number of computations, C, that each machine performs with the new algorithm is defined by

$$C = \min (C_1, C_2) \leq C_1, \tag{1}$$

where C_1 and C_2 are the computations that the machines would perform if they were operating independently. Let C_1 represent the computation to search the incorrect subset; then, moments of C are overbounded by the moments of the computation to decode the incorrect subset. Given that the first digit is decoded correctly, the moments of computation to decode the second digit are equal to the moments of C, etc.

3. Random Variable of Computation

We have not yet defined the term "computation." To do so, observe that for every path (a path is a consecutive sequence of any number of tree branches) which is acceptable, the following two branches must be examined. If we assume that both branches are examined simultaneously, then we can say that "one computation" is performed for each acceptable path.

Label paths in the tree with (n, s). The index s refers to the stage of the tree. The index n refers to the order of a path at the s^{th} stage, counted from the top, $1 \leq n \leq 2^s$.

Define the characteristic function $x_s(n)$ by

$$x_s(n) = \begin{cases} 1 & \text{if path } (n,s) \text{ is acceptable along its entire length} \\ 0 & \text{otherwise.} \end{cases}$$

Then, for each (n, s) in the incorrect subset for which $x_s(n) = 1$, one computation is necessary. Since the first branch is always examined, for C_1 we have

$$C_1 = 1 + \sum_{s=1}^{S-1} \sum_{n=1}^{2^{s-1}} x_s(n). \tag{2}$$

4. Bounds on Moments

Minkowski's inequality[2] plays a central role in the bounding of $\overline{C^p}$. If $x_1, \ldots, x_n$ are random variables, then the following theorem bounds moments of their sum as a function of their moments.

THEOREM 1:

$$\overline{(x_1+x_2+\ldots+x_n)^p}^{\,1/p} < \sum_{i=1}^{n} \overline{x_i^p}^{\,1/p}.$$

On the basis of this theorem, $\overline{C^p}$ has the bound

$$\overline{C^p}^{\,1/p} \leqslant \overline{C_1^p}^{\,1/p} \leqslant 1 + \sum_{s=1}^{S-1} \overline{\left\{\sum_{n=1}^{2^{s-1}} x_s(n)\right\}^p}^{\,1/p}. \tag{3}$$

Let us now bound $\overline{\left\{\sum\limits_n x_s(n)\right\}^p}$. The bound is written as a product.

$$\overline{\left\{\sum_{n=1}^{2^{s-1}} x_s(n)\right\}^p} = \sum_{n_1=1}^{2^{s-1}} \sum_{n_p=1}^{2^{s-1}} \overline{x_s(n_1)\ldots x_s(n_p)} \tag{4}$$

It should be noted that the indices $n_1, \ldots, n_p$ are dummy variables and that several may coincide. In particular, let us assume that the p-tuple $(n_1, \ldots, n_p) = \vec{n}_p$ has the t distinct elements $\theta_1, \ldots, \theta_t$. For example, the 6-tuple (12, 3, 7, 12, 6, 6) has 4 distinct elements, $\{\theta_i\} = \{3, 6, 7, 12\}$. Then, recognizing that $x_s(n)$ is either 0 or 1 so that $(x_s(n))^k = x_s(n)$, we have

$$\overline{x_s(n_1)\ldots x_s(n_p)} = \overline{x_s(\theta_1)\ldots x_s(\theta_t)}. \tag{5}$$

Because of Eq. 5, Eq. 4 will be evaluated by classifying p-tuples $\vec{n}_p$ in terms of their distinct elements $\theta_1, \ldots, \theta_t$ and summing over the various sets of distinct θ's. To perform this classification, observe that several p-tuples have the same set of θ's as elements. For example, the 5-tuples (1, 15, 4, 15, 6) and (6, 1, 1, 4, 15) have as elements 1, 4, 6, 15.

We ask now for a bound on the number f(t, p) (f(t, p) is independent of the particular set of t elements $\theta_1, \theta_2, \ldots, \theta_t$) of p-tuples, $\vec{n}_p$, with elements $\theta_1, \theta_2, \ldots \theta_t$. In every such p-tuple each element θ_i appears at least once; otherwise the p-tuple would contain less than t distinct elements. We now observe that this set of p-tuples is contained in the larger set of p-tuples constructed from the set of elements $\theta_1, \theta_2, \ldots, \theta_t$

where a particular $\vec{n}_p$ need not contain every θ_i. This larger set contains t^p p-tuples, since each of the p positions of a p-tuple may contain any one of t elements. Therefore,

$$f(t, p) \leq t^p. \tag{6}$$

We are now prepared to write Eq. 4 in terms of f(t, p). Using (5), we found that p-tuples should be classified according to their elements and that there are f(t, p) p-tuples that have a particular set $\theta_1, \theta_2, \ldots, \theta_t$ as elements. Therefore, the sum in (4) can be evaluated by summing over the various sets $\theta_1, \ldots, \theta_t$ with t fixed, multiplying by f(t, p), and summing on t. This verbal statement is presented formally in Eq. 7.

$$\sum_{n_1} \cdots \sum_{n_p} \overline{x_s(n_1) \ldots x_s(n_p)} = \sum_{t=1}^{\min(p, 2^{s-1})} f(t, p) \sum_{\substack{\text{all sets of } t \\ \text{distinct } \theta\text{'s}}} \overline{x_s(\theta_1) \ldots x_s(\theta_t)} \tag{7}$$

The upper limit on t reflects the fact that the number of distinct θ's cannot exceed the number of indices n_i, $1 \leq i \leq p$, or the number of values of any index n_i, $1 \leq n_i \leq 2^{s-1}$.

5. Topology of $\theta_1, \theta_2, \ldots, \theta_t$

As an aid in evaluating (7) we now examine the structure that the set of elements $\theta_1, \ldots, \theta_t$, at the s^{th} stage, gives to the tree code (see Fig. XIII-7). (Note that each

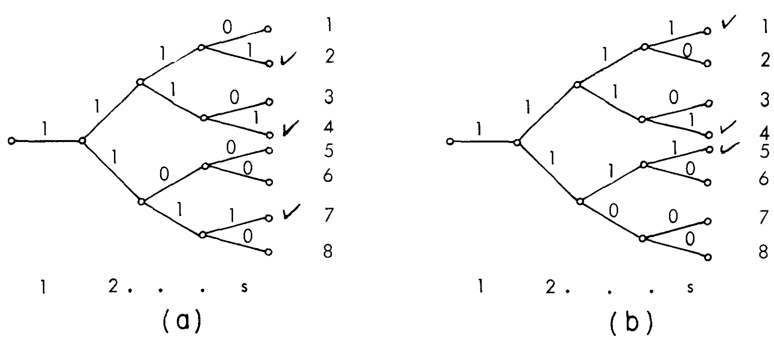

Fig. XIII-7. (a) Tree topology for $\{\theta_i\} = \{2, 4, 7\}$. (b) Tree topology for $\{\theta_i\} = \{1, 4, 5\}$.

θ_i defines a path in the tree.)

At the s^{th} stage of the tree, place a check next to each of the paths $\theta_1, \ldots, \theta_t$. Above every branch on a checked path place a 1 (see Fig. XIII-7a). Label all other paths with 0.

Let α_i equal the number of 1's at the i^{th} stage. For example, in Fig. XIII-7a, $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 3$, $\alpha_4 = 3$.

Now, to each distinct set of t elements $\{\theta_i\}$ there corresponds a unique tree array of 1's and 0's and a single s-tuple $\vec{\alpha} = (\alpha_1, \ldots, \alpha_s)$. For each s-tuple $\vec{\alpha}$, however, there are several sets $\{\theta_i\}$. For example, the set $\{\theta_i\} = \{2, 4, 7\}$ in Fig. XIII-7a and the set $\{\theta_i\} = \{1, 4, 5\}$ in Fig. XIII-7b both have $\vec{\alpha} = (1, 2, 3, 3)$.

Call $N_t(\vec{\alpha})$ the number of distinct sets $\{\theta_i\}$ with the assignment $\vec{\alpha}$. Clearly $N_t(\vec{\alpha}) = 1$ for $s = 1$.

LEMMA 1:

$$N_t(\vec{\alpha}) = \prod_{i=1}^{s-1} \binom{\alpha_i}{\alpha_{i+1} - \alpha_i} 2^{2\alpha_i - \alpha_{i+1}} \qquad s \geq 2$$

and

$$N_t(\vec{\alpha}) \leq 2\left(\frac{e}{4}\right)^t (s-1)^t \, 2^{\sum_{i=1}^{s} \alpha_i} \tag{8}$$

PROOF: We prove only the first statement. The number of arrangements of the α_{i+1} branches at the $(i+1)^{th}$ stage for each fixed arrangement of the α_i branches at the i^{th} stage is $\binom{\alpha_i}{\alpha_{i+1} - \alpha_i} 2^{2\alpha_i - \alpha_{i+1}}$. To see this, note that $\alpha_{i+1} - \alpha_i$ of the α_i branches at the i^{th} stage have two branchings and that $\alpha_i - (\alpha_{i+1} - \alpha_i)$ of the α_i branches have a single branching. The pairs can be arranged in $\binom{\alpha_i}{\alpha_{i+1} - \alpha_i}$ ways and for each arrangement of pairs the singles can be arranged in $2^{(2\alpha_i - \alpha_{i+1})}$ ways.

The product of the number of arrangements at the second, third, etc., stages is the total number of sets $\{\theta_i\}$ with the assignment $\vec{\alpha}$. Q.E.D.

It should be clear from the tree structure that the s-tuple $\vec{\alpha}$ must satisfy the following constraints: $\alpha_1 = 1$, $\alpha_s = t$, $1 \leq \alpha_i \leq \alpha_{i+1} \leq 2\alpha_i$.

6. Evaluation of $\overline{x_s(\theta_1) \ldots x_s(\theta_t)}$

The reason for the classification of sets $\{\theta_i\}$ according to their $\vec{\alpha}$ assignment is that the probability $\overline{x_s(\theta_1) \ldots x_s(\theta_t)}$ depends only on $\vec{\alpha}$. This statement will now be proved.

Assume that all averages are taken over the ensemble of source outputs, channel transitions, and the set of all tree codes. Over this ensemble each digit of the code is chosen independently and is equally likely to be chosen as 0 or 1, which implies that branches in different stages of the tree code are statistically independent. Also, successive channel transitions are statistically independent. Therefore, the probability,

$\overline{x_s(\theta_1) \dots x_s(\theta_t)}$, that the paths $\theta_1, \dots, \theta_t$ with assignment $\vec{a}$ have every branch acceptable can be written as a product.

$$\overline{x_s(\theta_1) \dots x_s(\theta_t)} = \prod_{i=1}^{s} \Pr(\text{all } a_i \text{ branches at the } i^{\text{th}} \text{ stage are acceptable})$$

LEMMA 2: If paths $\theta_1 \dots \theta_t$ have the assignment $\vec{a}$, then

$$\overline{x_s(\theta_1) \dots x_s(\theta_t)} = 2^{-\sum_{i=1}^{s} a_i \frac{R_{a_i}}{R}} \leq 2^{-\frac{R_p}{R} \sum_{i=1}^{s} a_i}, \tag{10}$$

where

$$R_k = -\frac{1}{k} \log_2 \left(g_1 + \frac{g_o}{2^k}\right). \tag{11}$$

PROOF:

$$\overline{x_s(\theta_1) \dots x_s(\theta_t)} = \prod_{i=1}^{s} \sum_{k_i=0}^{\ell} \Pr(\text{all } a_i \text{ branches acceptable}/k_i \text{ erasures}) \Pr(k_i \text{ erasures})$$

where $\Pr(k_i \text{ erasures}) = \binom{\ell}{k_i} g_1^{k_i} g_o^{\ell-k_i}$ and ℓ is the number of digits on a branch. Over the ensemble of source outputs and tree codes each digit chosen for a code is equally likely to be 0 or 1. Thus, $\Pr(\text{all } a_i \text{ branches acceptable}/k_i \text{ erasures}) = \left(1/2^{a_i}\right)^{\ell-k_i}$. Defining

$$R_k = \frac{-\log_2 \left(g_1 + \frac{g_o}{2^k}\right)}{k}$$

and noting that $0 \leq R_{k+1} < R_k$, we verify both hypotheses. Q.E.D.

7. Bounds on $\overline{C^p}$

Combining (7) and (8) and (10), we have

$$\overline{\left\{\sum_{n=1}^{2^{s-1}} x_s(n)\right\}^p} = \sum_{t=1}^{\min(p, 2^{s-1})} f(t, p) \sum_{\vec{a}} N_t(\vec{a})\, 2^{-\sum_{i=1}^{s} a_i \frac{R_{a_i}}{R}}, \tag{12}$$

where the summation on $\vec{a}$ is performed subject to the constraints $a_1 = 1$, $a_s = t$,

$1 \leq a_i \leq a_{i+1} \leq 2a_i$. The summation on $\vec{a}$ with the weighting $N_t(\vec{a})$ is equivalent to the summation over all distinct sets of t θ's in (7).

The bounds on the various quantities which were found earlier are now employed to bound the p^{th} moment of the sum above and to bound $\overline{C^p}$. The results are stated without proof in Theorem 2 but are easily obtained from earlier bounds.

THEOREM 2: The p^{th} moment of the random variable, C, the number of computations to decode a single source digit, when averaged over the ensemble of all tree codes, has the overbounds

$$\overline{C^p} \leq \begin{cases} (2^S)^p & \\ 2(S^3p)^p & R \leq R_p \\ 2\left\{\dfrac{3p}{\left[1-2^{-\frac{1}{p}\left(\frac{R_p}{R}-1\right)}\right]^3}\right\}^p & R \leq R_p, \end{cases} \tag{13}$$

where $R_p \triangleq 1 - \frac{1}{p}\log_2 (g_1 2^p + g_o)$. The best bound is a function of p, S, and rate R.

The rates R_p are related to a bound on the best performance of the Binary Erasure

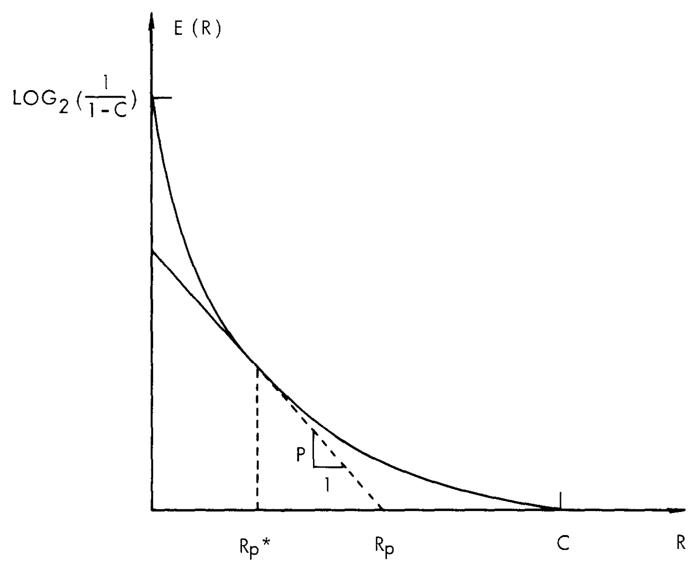

Fig. XIII-8. Construction for R_p.

Channel as follows: It has been shown[3] that the lower bound to the probability of error for maximum-likelihood decoding of block codes has an exponent E(R), where

$$E(R) = (1-R) \log_2 \left(\frac{1-R}{1-C}\right) + R \log_2 \left(\frac{R}{C}\right)$$

and $C = g_o$ = channel capacity. The rates R_p are found by intersecting with the rate axis a line of slope $-p$ tangent to $E(R)$ (see Fig. XIII-8). The rate R_p^* is the point of tangency of the straight line.

The rates R_p are also related to the exponent on the average probability of error with "list decoding."[4,5] To "list-decode" a block code a list is made of the p a posteriori most probable code words; the decoder presents as output this set; an error is said to occur if the decoded set of p words does not contain the transmitted word; the exponent on this probability of error is equal to $E(R)$ for $R \geq R_p^*$ and $p[R_p-R]$ for $R \leq R_p^*$ which is the straight-line extension of $E(R)$ from $E(R_p^*)$ at $R = R_p^*$ (see Fig. XIII-8).

8. Bound on the Distribution of Computation

We now bound the distribution of computation by using Lemma 3 (which is a form of Tchebycheff's inequality) and Theorem 2.

LEMMA 3: For a positive, integer-valued random variable x,

$$\Pr[x \geq N] \leq \frac{\overline{x^k}}{N^k}. \tag{14}$$

PROOF:

$$\overline{x^k} = \sum_{x=1} x^k p(x) \geq \sum_{x=N} x^k p(x) \geq N^k \Pr[x \geq N]. \qquad \text{Q.E.D.}$$

Then, for $k = p$ and $R < R_p$

$$\Pr[x \geq N] \leq 2 \left\{ \frac{3p}{\left[1 - 2^{-(1/p)((R_p/R)-1)}\right]^3} \right\}^p \frac{1}{N^p} \tag{15}$$

is an example of a bound on the distribution.

This is a bound on the average probability that more than N units of computation are necessary to decode the first source digit. It is an average over the set of all tree codes. Over this ensemble the distribution of computation is a positive random variable. As such, more than $\left(1 - \frac{1}{a}\right)$ of the codes picked from this ensemble will have a distribution of computation $\Pr^*[x \geq N]$ with

$$\Pr^*[x \geq N] \leq 2a \left\{ \frac{3p}{\left[1 - 2^{-(1/p)((R_p/R)-1)}\right]^3} \right\}^p, \frac{1}{N^p} \qquad R < R_p. \tag{16}$$

It should be observed that bounds have been obtained only for the distribution of a single random variable of computation. The dependence of the computation on successive source digits has not been studied.

The continuing study of sequential decoding has as its object to extend the successful analysis of the Wozencraft algorithm to the Fano algorithm. Also, an effort is being made to extend the results from channels uniform at the output to a larger class of channels. The buffer problem is also being examined.

J. E. Savage

References

1. For a description of the algorithm, see B. Reiffen, Sequential Encoding and Decoding for the Discrete Memoryless Channel, Technical Report 374, Research Laboratory of Electronics, M.I.T., August 12, 1960.

2. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities (Cambridge University Press, New York, 1959).

3. R. M. Fano, Transmission of Information (The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961), Chapter 9.

4. P. Elias, List Decoding for Noisy Channels, Technical Report 335, Research Laboratory of Electronics, M.I.T., September 20, 1957.

5. J. M. Wozencraft, List decoding, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., January 15, 1958, p. 90.